

Blending/Reblending Library
Services: Supporting
Reproducible Science

Harrison Dekker - @vagrantscholar
University of Rhode Island

Reproducibility crisis

“A methodological crisis in science in which scientists have found that the results of many scientific experiments are difficult or impossible to replicate on subsequent investigation, either by independent researchers or by the original researchers themselves.”

– Wikipedia

Communicating computational results

```
In [1]: import pywikibot

In [6]: site = pywikibot.Site('wikidata', 'wikidata')

In [9]: site
Out[9]: DataSite('wikidata', 'wikidata')

In [13]: item = pywikibot.ItemPage(site, 'Q4115189')

In [14]: item
Out[14]: ItemPage('Q4115189')

In [15]: item_dict = item.get()

In [16]: item_dict
Out[16]: {'aliases': {'ar': ['ساندوك الحريمية', 'ساندوك التجريبية'],
  'de': ['Spielweine', 'Sandbox'],
  'de-at': ['Sandkasten', 'Spielplatz'],
  'en': ['SB',
  'Property test',
  'test',
  'wikidata SandboxItem',
  'wikidata 000'],
  'it': ['sandbox di Wikidata'],
  'ja': ['サンドボックス', '練習用<!-->', '練習用項目'],
  'nl': ['Wikidata-speelruimte'],
  'pt-br': ['item para testes', 'teste', 'testes', 'test', 'página de testes'],
  'ru': ['тест', 'тест2'],
  'zh-hans': ['维基数据沙盒', '维基数据测试']},
  'claims': {'P1110': ['qpywikibot.page.Claim at 0x77543b30390d'],
  'P1132': ['qpywikibot.page.Claim at 0x77543b30658d'],
  'P1302': ['qpywikibot.page.Claim at 0x77543b30619d',
  qpywikibot.page.Claim at 0x77543b3062bd,
  qpywikibot.page.Claim at 0x77543b3066bd],
  'P1346': ['qpywikibot.page.Claim at 0x77543b303f6d'],
  'P1350': ['qpywikibot.page.Claim at 0x77543b303e8d'],
  'P1351': ['qpywikibot.page.Claim at 0x77543b30a58d'],
  'P1355': ['qpywikibot.page.Claim at 0x77543b30a4fd'],
  'P1356': ['qpywikibot.page.Claim at 0x77543b30789d'],
  'P18': ['qpywikibot.page.Claim at 0x77543b309f1d'],
  'P1623': ['qpywikibot.page.Claim at 0x77543b30868d'],
  'P2047': ['qpywikibot.page.Claim at 0x77543b307fd'],
  'P2030': ['qpywikibot.page.Claim at 0x77543b3066d'],
  'P27': ['qpywikibot.page.Claim at 0x77543b3080d'],
  'P279': ['qpywikibot.page.Claim at 0x77543b3088d'],
  'P91': ['qpywikibot.page.Claim at 0x77543b30827b',
  qpywikibot.page.Claim at 0x77543b3083d],
  'P426': ['qpywikibot.page.Claim at 0x77543b306cfd'],
  'P488': ['qpywikibot.page.Claim at 0x77543b306a8d']},
```

Modern data analysis typically involves dozens, if not hundreds of steps, each of which can be performed by numerous algorithms that are nominally identical but differ in detail, and each of which involves at least some ad hoc choices. If researchers do not make their code available, there is little hope of ever knowing what was done to the data, much less assessing whether it was the right thing to do.

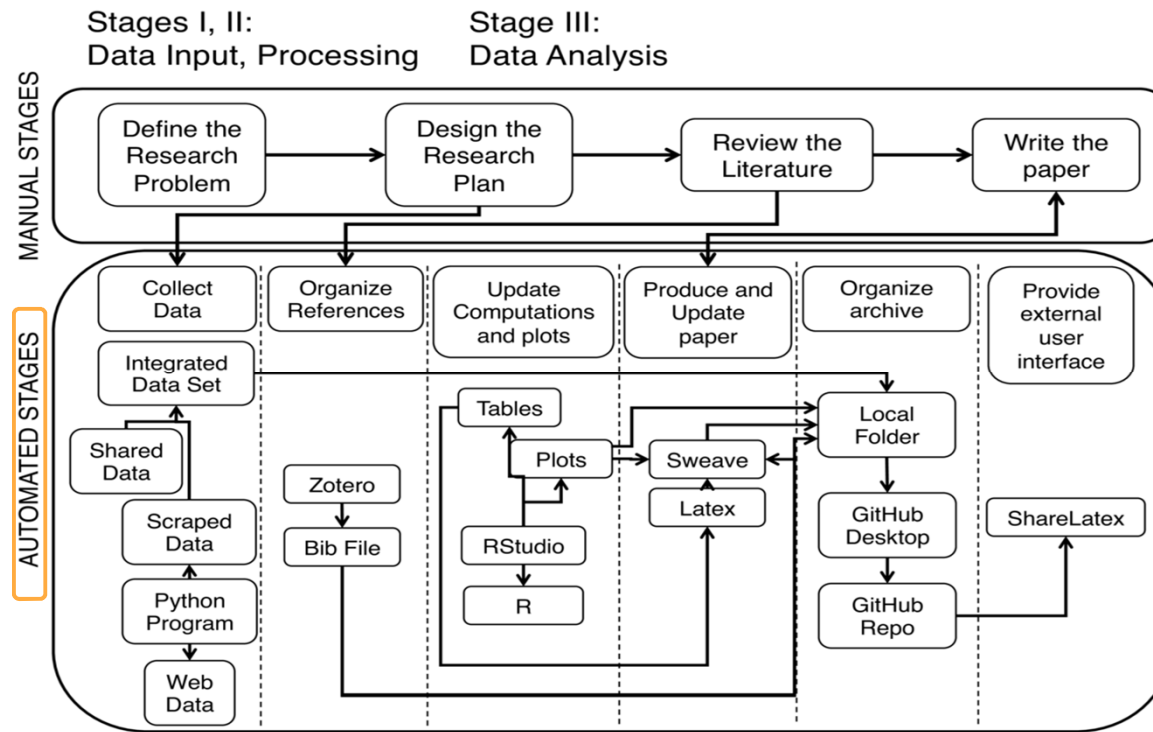
Stark, 2018

[Screenshot](#) by By Tobias1984 (Own work) is licensed under [CC BY-SA 4.0](#)

General workflow model



Detailed workflow example



Magallanes, 2018

Workflow skills and tools

Skill type	Description	Tools
Literate computing	Enable writing self-contained documents combining text and code	Rstudio : Markdown : LaTeX : Jupyter
Version control	Track file changes over time. Revert to earlier versions. Branch/fork	Git : GitHub : BitBucket : Open Science Framework
Tracking provenance	Capture complex workflows involving multiple research objects/tools	VisTrails : Kepler : Taverna
Automation	Automate workflows using time-tested and ubiquitous command line tools	Unix command line : shell scripts : make
Virtual environments	Capture complex computation environments and configurations	VirtualBox : VMWare : Docker

<https://ropensci.github.io/reproducibility-guide/sections/introduction/>, 2018

More information

Training:

Data/Software Carpentry -- <https://carpentries.org/>

Library Carpentry -- <https://librarycarpentry.github.io/>

Case Studies:

Kitzes, J., Turek, D., & Deniz, F. (Eds.). (2018). *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences*. Oakland, CA: University of California Press.

(A free [pre-print edition](#) is available)

Teaching materials:

Project TIER -- <https://www.projecttier.org/>

ROpenSci -- <https://ropensci.github.io/reproducibility-guide/>

BITSS -- <https://www.bitss.org/resources/>